

Identification of Disease Genes

This mini-course focuses on the identification of a disease gene using NCBI's human genome assembly. The reference human genome assembly along with integrated maps, literature, and expression information comprises a powerful discovery system for exploring candidate human disease genes.

Problem: A laboratory has generated an EST library from a hemochromatosis patient and wants to identify the gene(s) causing the phenotype.

We will follow these steps to solve the problem:

1. Compare ESTs to the human genome (using BLAST).
2. Identify the gene(s) aligning the ESTs and download their sequences (using MapViewer).
3. Identify whether the ESTs contain any known SNPs (using dbSNP).
4. Determine whether a mutant form of the gene causes a phenotype (using OMIM).

A web page

(<http://www.ncbi.nlm.nih.gov/Class/minicourses/diseasegene.html>) describes in detail how to perform these steps.

The following handout includes the screen shots of the exercise.

Course developed by: Medha Bhagwat (bhagwat@ncbi.nlm.nih.gov)

Instructor: Wayne Matten (matten@ncbi.nlm.nih.gov)

Problem 1:

A laboratory has generated an EST library from a hemochromatosis patient and wants to identify the gene(s) causing the phenotype.

Outline:

We will follow these steps to solve the problem:

1. Compare ESTs from a hemochromatosis patient to the human genome (using BLAST).
2. Identify the gene(s) aligning the ESTs and download their sequences (using Map Viewer).
3. Identify whether the ESTs contain any known nucleotide variations (single nucleotide polymorphisms) (using dbSNP).
4. Determine whether a mutant form of the gene is known to cause a phenotype (using OMIM).

Step 1. Compare ESTs to the human genome (using BLAST):

One way to identify the genes expressing the ESTs is to compare their sequences using BLAST with the human genome assembly and the genes annotated on it. To access the specialized BLAST page for searching against the human genome assembly, click on

[BLAST \(human genome\)](#)

Paste the EST sequence provided below in the query box of the BLAST page and start the search by clicking on the “Begin Search” button.

Query EST Sequence:

```
TGCCTCCTTTGGTGAAAGGTGACACATCATGTGACCTCTTCAG  
ATCAACCATGAAAGTGGCTGAAGGATAAGCAGCCAATGGATG  
CTACCA GGGCTGGATAACCTTGGCTGTACCCCTGGGGAAG  
CCCTCATTTGTGATCTGGG
```

Name the chromosome and the contig that we get as a BLAST hit. Is the EST sequence 100% identical to the genomic sequence? Note the nucleotide difference between the two sequences. Paste your results in the window below.

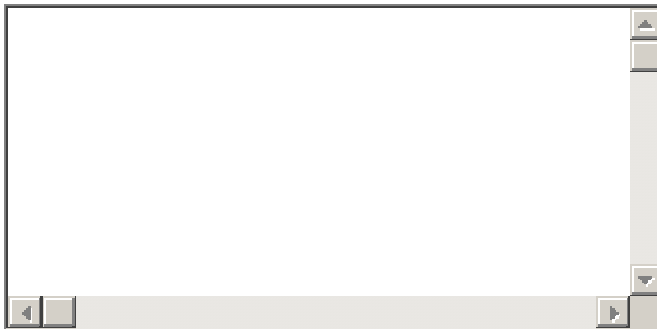
Results of BLAST against the human genome

Step 2. Identify the gene(s) expressing the ESTs and download their sequences:

To visualize the BLAST hit on the genome using Map Viewer, click on the "Genome View" button at the top of the results page, then on the Map element "NT_007592". Currently, 4 maps should be displayed (Contig, Model, RNA and Gene_seq). Zoom out 2 or 4 times by clicking on right most contig map and selecting the appropriate option.

The BLAST hit, indicated by the red bar, is in the region of one of the exons of the HFE gene annotated on the human genome. Make the Gene_seq map a master map by clicking on the arrow at the top of the map. Display the entire HFE gene sequence by clicking on the "dl" link and then on "Display". Copy the sequence and paste it in the area provided below. We will use it later to obtain the exon-intron structure. You can adjust the nucleotide locations to download the upstream or downstream sequence by using the "adjust by" and "Change Region/Strand" option.

HFE gene sequence



Step 3. Determine whether the ESTs contain known SNPs:

Go back to the Map Viewer report. Click on the Maps and Options link. Remove all the maps except the Gene_seq map by selecting the map under the Maps Displayed menu and clicking on Remove. Now add the variation map from the Available maps menu (by selecting the map and clicking on Add). Make the Variation map as the master map by selecting it and clicking the Make Master/Move to Bottom option. Then click on Apply. Now two maps are displayed, Variation (it's the rightmost and master map) and Gene_seq. The master map provides detailed information for the map features, in this case SNPs. ". (The Mini-Course Map Viewer Quick Start describes the usage of the Map Viewer in detail.) Zoom in on the blast hit area (red bar). There are two SNPs in the area, one of them is rs1800562. Click on the link for the SNP. There

is an A/G SNP is at the nucleotide position 16951392 on the contig NT_007592 as mentioned under Fasta sequence and Integrated maps. Is this the same nucleotide variation found in the BLAST result in Step 1? Please note that the SNP results in the Cysteine 282 Tyrosine mutation for the longest protein (expressed by the mRNA NM_000410) as reported under GeneView.

Step 4. Determine whether the mutant HFE gene causes a phenotype:

Go back to the Map Viewer report. Make the Gene_seq map as the master map. Select the link to the OMIM database. It takes us to the OMIM report for the HFE gene that details how mutations in the HFE gene are associated with a phenotype, hemochromatosis. Click on the Allelic Variant “View list” to get information about mutant proteins from patients. Is Cys282Tyr variant mentioned in the list? Which phenotype does it cause?

Summary:

This mini-course describes steps to identify the gene expressing the ESTs obtained from a hemochromatosis patient, download the gene sequence, identify known SNPs in the gene and find SNP-associated phenotypes.

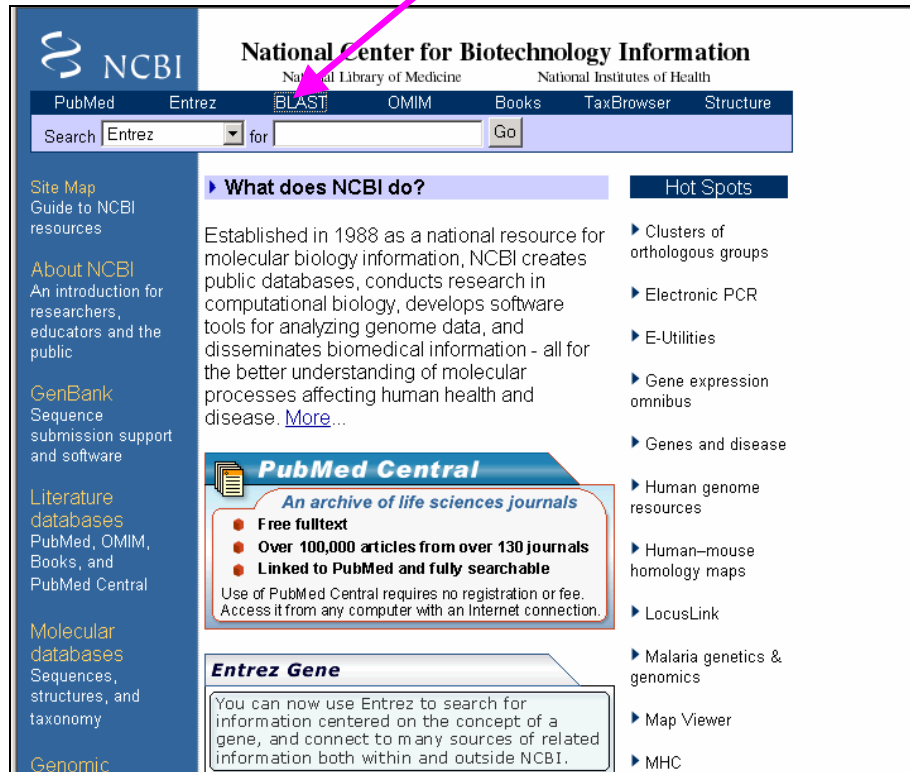
Step 1: The query EST sequence was found to align contig NT_007592.14 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 16951392 on the contig).

Step 2: The query EST was found to be expressed by the HFE gene.

Step 3: The query EST sequence contains a known SNP (G/A with respect to the nucleotide 16951392 on contig NT_007592.14).

Step 4: Mutations in the HFE gene are associated with hemochromatosis.

Step 1: Compare ESTs against the human genome



NCBI National Center for Biotechnology Information
National Library of Medicine National Institutes of Health

PubMed Entrez **BLAST** OMIM Books TaxBrowser Structure

Search for

Site Map
Guide to NCBI resources

About NCBI
An introduction for researchers, educators and the public

GenBank
Sequence submission support and software

Literature databases
PubMed, OMIM, Books, and PubMed Central

Molecular databases
Sequences, structures, and taxonomy

Genomic

What does NCBI do?
Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More...](#)

Hot Spots

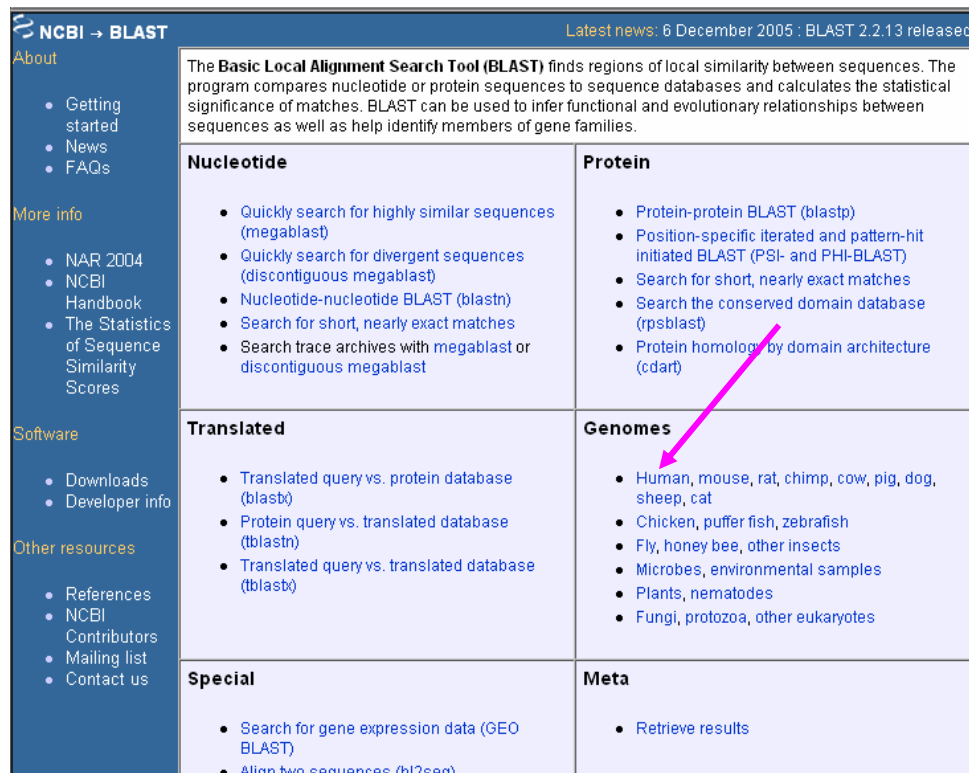
- Clusters of orthologous groups
- Electronic PCR
- E-Utilities
- Gene expression omnibus
- Genes and disease
- Human genome resources
- Human-mouse homology maps
- LocusLink
- Malaria genetics & genomics
- Map Viewer
- MHC

PubMed Central
An archive of life sciences journals

- Free fulltext
- Over 100,000 articles from over 130 journals
- Linked to PubMed and fully searchable

Use of PubMed Central requires no registration or fee. Access it from any computer with an Internet connection.

Entrez Gene
You can now use Entrez to search for information centered on the concept of a gene, and connect to many sources of related information both within and outside NCBI.



NCBI → BLAST Latest news: 6 December 2005 : BLAST 2.2.13 released

About

- Getting started
- News
- FAQs

More info

- NAR 2004
- NCBI Handbook
- The Statistics of Sequence Similarity Scores

Software

- Downloads
- Developer info

Other resources

- References
- NCBI Contributors
- Mailing list
- Contact us

The **Basic Local Alignment Search Tool (BLAST)** finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

Nucleotide	Protein
<ul style="list-style-type: none"> Quickly search for highly similar sequences (megablast) Quickly search for divergent sequences (discontiguous megablast) Nucleotide-nucleotide BLAST (blastn) Search for short, nearly exact matches Search trace archives with megablast or discontiguous megablast 	<ul style="list-style-type: none"> Protein-protein BLAST (blastp) Position-specific iterated and pattern-hit initiated BLAST (PSI- and PHI-BLAST) Search for short, nearly exact matches Search the conserved domain database (rpsblast) Protein homology by domain architecture (cdart)
Translated <ul style="list-style-type: none"> Translated query vs. protein database (blastx) Protein query vs. translated database (tblastn) Translated query vs. translated database (tblastx) 	Genomes <ul style="list-style-type: none"> Human, mouse, rat, chimp, cow, pig, dog, sheep, cat Chicken, puffer fish, zebrafish Fly, honey bee, other insects Microbes, environmental samples Plants, nematodes Fungi, protozoa, other eukaryotes
Special <ul style="list-style-type: none"> Search for gene expression data (GEO BLAST) Align two sequences (bl2seq) 	Meta <ul style="list-style-type: none"> Retrieve results

BLAST Human Sequences. - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Reload Home Search Favorites

Address <http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9606> Go Links

NCBI Home > Genomic Biology > Human Genome Resources > BLAST

Search Map Viewer Go Clear

BLAST
overview
FAQs
news
manual
references
Retrieve results
Genome Project

BLAST Human Sequences.

☒ Enter an accession, gi, or a sequence in FASTA format:

```
TGCCTCCTTTGGTGAAAGGTGACACATCATGTGACCTCTTCAGTGACCCTCTACGGTGTGGGCC
TTGAACTACTACCCCAAGAAC
ATCACCATGAAGTGGCTGAAGGATAAGCAGCCAATGGATGCCAAGGAGTTTGAACCTAAAGACGT
ATTGCCCAATGGGGATGGGAC
CTACCAGGGCTGGATAACCTTGGCTGTACCCCTGGGGAAGAGCAGATATACGTACCAGGTGG
AGCACCAGGCCTGGATCAGC
```

☐ Or, choose a file to upload

Browse...

Database:

- genome (all assemblies)
- genome (all assemblies)
- genome (reference only)
- HTGS
- RefSeq protein
- Non-RefSeq RNA
- Non-RefSeq protein
- Build RNA
- Build protein
- Ab initio RNA
- Ab initio protein

sequences

Alignments

Advanced options:

Begin Search Clear Input

Local intranet

NCBI *formatting* **BLAST**

Nucleotide Protein Translations Retrieve results for an RID

Your request has been successfully submitted and put into the Blast Queue.

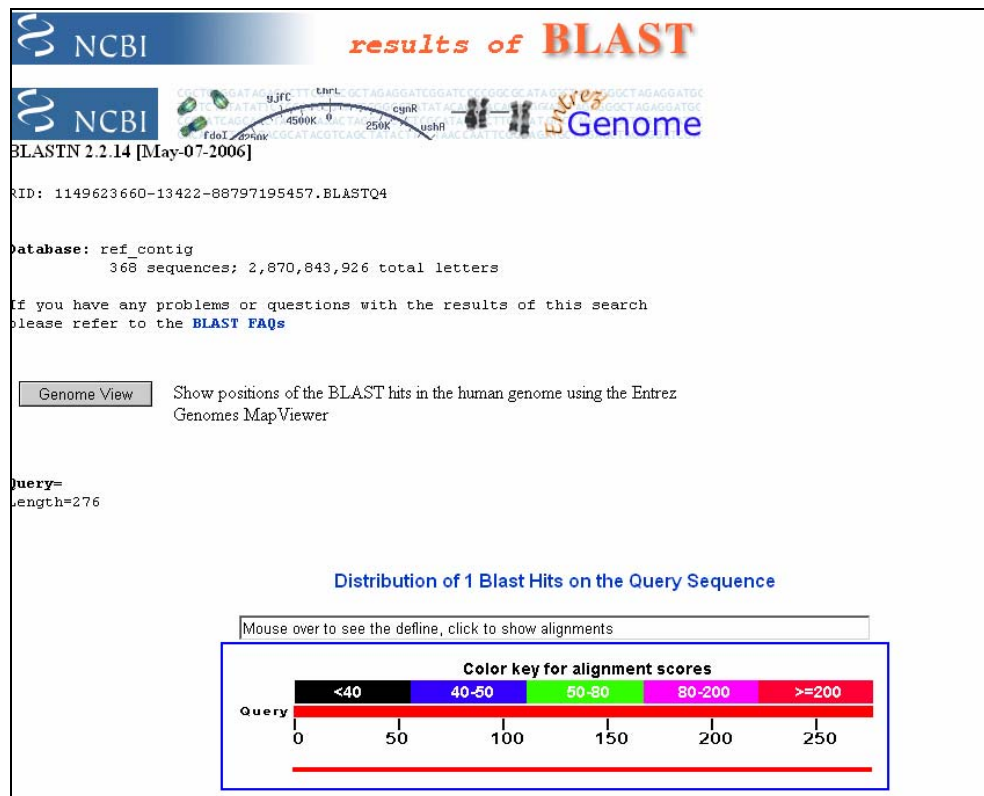
Query = (276 letters)

The request ID is

Format! or **Reset all**

The results are estimated to be ready in 32 seconds but may be done sooner.

Please press "FORMAT!" when you wish to check your results. You may change the formatting options for your result via the form below and press "FORMAT!" again. You may also request results of a different search by entering any other valid request ID to see other recent jobs.



Alignments

>[ref|NT_007592.14|Hs6_7749](#) D Homo sapiens chromosome 6 genomic contig
Length=48945890

Features in this part of subject sequence:

[hemochromatosis protein isoform 11 precursor](#)
[hemochromatosis protein isoform 10 precursor](#)

Score = 505 bits (273), Expect = 6e-141
Identities = 275/276 (99%), Gaps = 0/276 (0%)
Strand=Plus/Plus

Query	1	TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCTACGGTGTC	60
Sbjct	16951164	TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCTACGGTGTC	16951223
Query	61	GGGCCTTGAACCTACTACCCCAAGAACATCACCATGAAGTGGCTGAAGGATAAGCAGCCAA	120
Sbjct	16951224	GGGCCTTGAACCTACTACCCCAAGAACATCACCATGAAGTGGCTGAAGGATAAGCAGCCAA	16951283
Query	121	TGGATGCCAAGGAGTTTGAACCTAAAGACGTATTGCCCAATGGGGATGGGACCTACCAGG	180
Sbjct	16951284	TGGATGCCAAGGAGTTTGAACCTAAAGACGTATTGCCCAATGGGGATGGGACCTACCAGG	16951343
Query	181	GCTGGATAACCTTGGCTGTACCCCTGGGGAAGAGCAGAGATATCGTACCAGSTGGAGC	240
Sbjct	16951344	GCTGGATAACCTTGGCTGTACCCCTGGGGAAGAGCAGAGATATCGTACCAGSTGGAGC	16951403
Query	241	ACCCAGGCCTGGATCAGCCCTCATTGTGATCTGGG	276
Sbjct	16951404	ACCCAGGCCTGGATCAGCCCTCATTGTGATCTGGG	16951439

Result: The EST sequence is aligned to the contig NT_007592.14 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 16951392 on the contig).

Step 2: Identify the gene(s) expressing the ESTs and download their sequences

NCBI BLASTN 2.2.14 [May-07-2006]

RID: 1149623660-13422-88797195457.BLASTQ4

Database: ref_contig
368 sequences; 2,870,843,926 total letters

If you have any problems or questions with the results of this search please refer to the [BLAST FAQs](#)

[Genome View](#) Show positions of the BLAST hits in the human genome using the Entrez Genomes MapViewer

Query=
Length=276

NCBI Map Viewer

PubMed Nucleotide Protein Genome Gene Structure PopSet Taxonomy Help

Search for on chromosome(s) assembly All

☐ Show related entries

Homo sapiens (human) genome view [BLAST search the human genome](#)

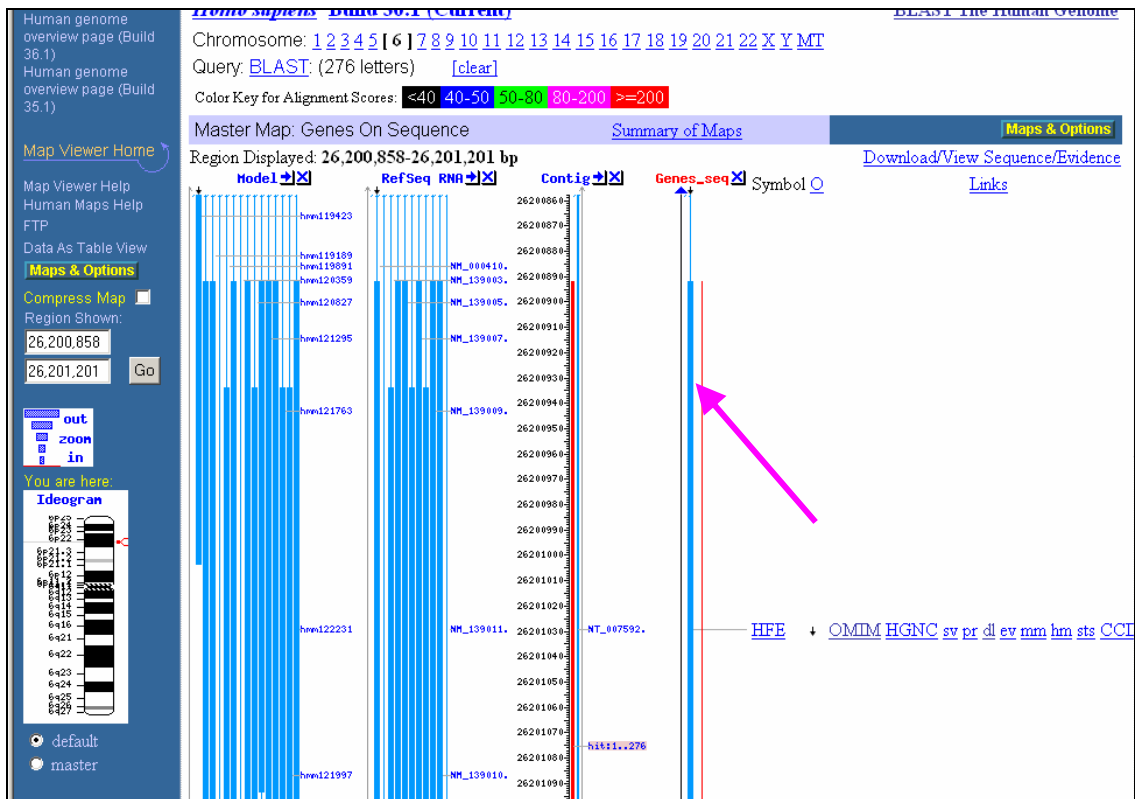
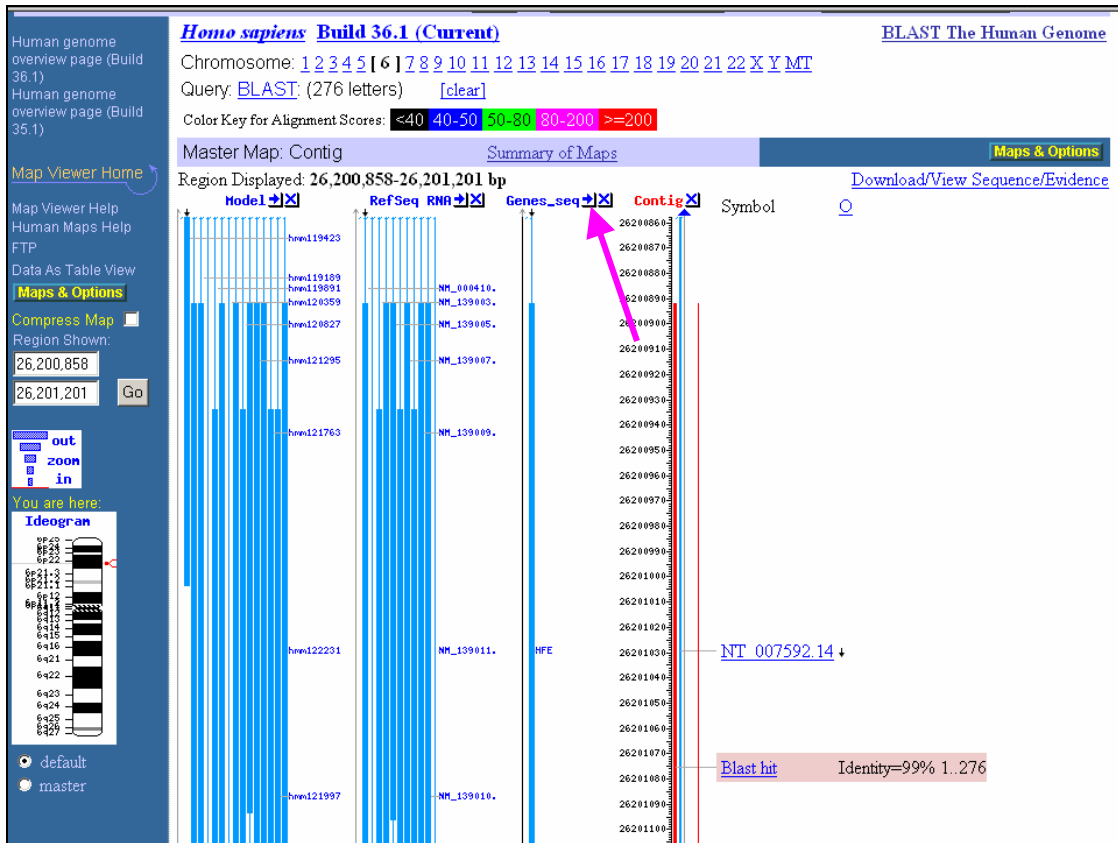
[Build 36.1 statistics](#) [Switch to previous build](#)

Color key for scores: < 40 40-50 50-80 80-200 >= 200

[Back to BLAST alignments page](#)

BLAST search results: 1 BLAST hit found (Request ID "1150133069-18789-55767668322.BLASTQ4").

Chr	Map element	Type	Hits	Score	E value
6	NT_007592	CONTIG	1	505	6e-141



Human genome overview page (Build 36.1)
 Human genome overview page (Build 35.1)

[Map Viewer Home](#)

Map Viewer Help
 Human Maps Help
 FTP
 Data As Table View
[Maps & Options](#)

Compress Map ☐

Region Shown:
 26,200,858
 26,201,201

You are here:
 Ideogram

☐ default
☐ master

[Homo sapiens Build 36.1 \(Current\)](#) [BLAST The Human Genome](#)

Chromosome: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [X](#) [Y](#) [MT](#)

Query: [BLAST](#): (276 letters)

Color Key for Alignment Scores: <40 40-50 50-80 80-200 >=200

Master Map: Genes On Sequence [Summary of Maps](#) [Maps & Options](#)

Region Displayed: 26,200,858-26,201,201 bp

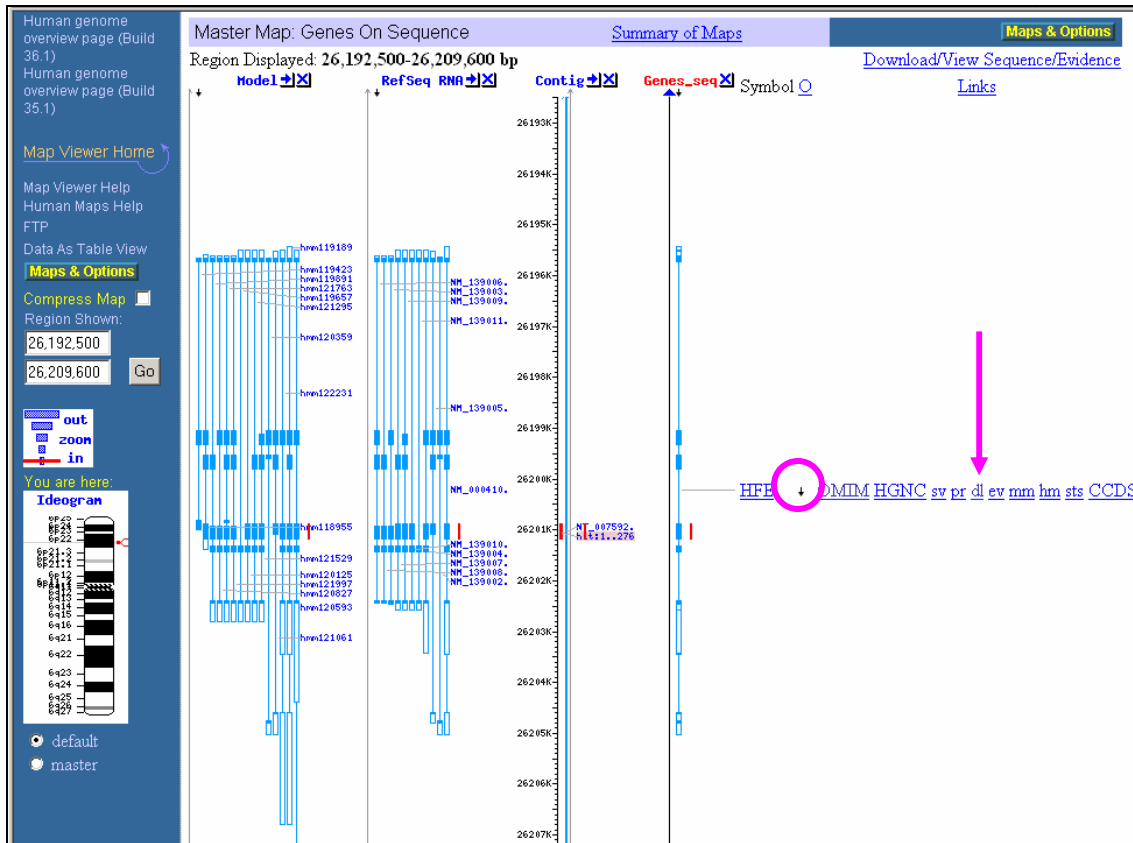
[Model](#) [RefSeq](#) [RNA](#) [Contig](#) [Genes_seq](#) [Symbol](#)

[Download/View Sequence/Evidence](#) [Links](#)

MapView

- Recenter
- Zoom in x2
- Zoom in x4
- Zoom in x8
- Zoom out x2
- Show 1 M
- Show 1 K
- Show 10 K
- Show 100 K
- Show Sequence

[OMIM](#) [HGNC](#) [sv](#) [pr](#) [di](#) [ev](#) [mm](#) [hm](#) [sts](#) [CC](#)



[Homo sapiens](#) (Build 36.1)
Region to retrieve (in chromosome coordinates):
Chromosome: Strand:
From: adjust by:
to: adjust by: [Change Region/Strand](#)
Sequence Format:

This chromosome region corresponds to the contig region(s):

Contig	start	stop	strand	
NT_007592.14	16945699	16955310	+	Display Save to Disk View Evidence ModelMaker

Range: from **16945699** to **16955310** [Show whole sequence](#) ☐ Reverse complemented strand [Refresh](#)

☐ **1:** [NT_007592](#). [Reports](#) Homo sapiens chro...[gi:51465675]

```
>ref|NT_007592.14|Hs6_7749:16945699-16955310 Homo sapiens chromosome 6 genomic contig
GGGGACACTGGATCACCTAGTGTTCACAAGCAGGTACCTTCTGCTGTAGGAGAGAGAACTAAAGTTC
TGAAAGACCTGTTGCTTTTACCAGGAAGTTTACTGGGCATCTCCTGAGCCTAGGCAATAGCTGTAGGG
TGACTTCTGGAGCCATCCCCGTTTCCCCGCCCCCAAAAAGAGCGGAGATTTAACGGGGACGTGCGGCCA
GAGCTGGGGAATGGGCCCCGAGCCAGGCCGGCGCTTCTCCTCCTGATGCTTTTGCAGACCGCGGTCTT
GCAGGGGCGCTTGTGCGTGAGTCCGAGGGCTGCGGGCGAACTAGGGGCGCGCGGGGGTGGAAAAATCG
AAACTAGCTTTTCTTTGCGCTTGGGAGTTTGCTAACTTTGGAGGACCTGCTCAACCCATATCCGCAAGCC
CCTCTCCCTACTTTCTGCGTCCAGACCCCGTGAGGGAGTGCCTACCACTGAACTGCAGATAGGGGTCCCT
CGCCCCAGGACCTGCCCCCTCCCCCGGCTGTCCCGGCTCTGCGGAGTGACTTTTGGAAACCGCCACTCCC
TTCCCCCACTAGAAATGCTTTTAAATAAATCTCGTAGTTCCTCACTTGAGCTGAGCTAAGCCTGGGGCTC
CTTGAACCTGGAACTCGGGTTTATTTCCAATGTCAGCTGTGCAGTTTTTTCCCAAGTCATCTCCAAACAG
GAAGTTCTTCCCTGAGTGCTTGCCGAGAAAGGCTGAGCAAAACCCACAGCAGGATCCGCACGGGGTTTCCAC
CTCAGAACGAATGCGTTGGGCGGTGGGGGCGCGAAAAGAGTGGCGTTGGGGATCTGAATTCTTCAACCATTC
CACCCACTTTTGGTGAGACCTGGGGTGAGGTCTCTAGGGTGGGAGGCTCCTGAGAGAGGCCTACCTCGG
GCCTTTCCCCCACTCTTGGCAATTGTTCTTTTGCTGGAAAATTAAGTATATGTTAGTTTGAACGTTTGA
ACTGAACAATTCTCTTTTCGGCTAGGCTTTATTGATTGCAATGTGCTGTGTAATTAAGAGGCCCTCTCTA
CAAAGTACTGATAATGAACATGTAAGCAATGCACTCACTTCTAAAGTTACATTTCATATCTGATCTTATTG
ATTTTCACTAGGCATAGGGAGGTAGGAGCTAATAATACGTTTATTTTACTAGAAAGTTAACTGGAATTCAG
ATTATATAACTCTTTTCAGGTTACAAAAGAACATAAATAATCTGGTTTTCTGATGTTATTTCAAGTACTAC
AGCTGCTTCTAATCTTAGTTGACAGTGATTTTGGCCTGTAGTGAGCAGAGTGTCTGTGGGTACACGCG
CGGCCTCAGCACAGCACTTTGAGTTTGGTACTACGTGTATCCACATTTTACACATGACAAAGAATGAGGC
ATGGCACGGCCTGCTTCTGGCAAAATTTATTCAATGGTACACTGGGCTTTGGTGGCAGAGCTCATGTCTC
```

Result: The query EST is expressed by the HFE gene.



Organism: Homo sapiens [Help](#)

Chromosome: Region Shown:

Available Maps: **Maps Displayed (left to right):**

Org: Assembly:

–Sequence Maps–

- Ab initio
- Assembly
- BES_Clone
- Clone
- NCI_Clone
- Contig
- Component
- CpG Island

- ☐ Ab initio
- ☐ Transcript (RNA)
- ☐ Gene
- ☒ [R] Contig

([R] before map means 'ruler set')

More Options:

☐ Show Connections ☒ Verbose Mode

Compress Map: Auto Compress if > px

Page Length:

Thumbnail View: ☒ default (ideogram) ☐ master

Organism: Homo sapiens [Help](#)

Chromosome: Region Shown:

Available Maps: **Maps Displayed (left to right):**

Org: Assembly:

Gga_UniGene

Gga_EST

Variation

–Cytogenic maps–

Ideogram

FISH Clone

NCI FISH Clone

Gene_Cytogenetic

Mitelman Breakpoint

- ☐ Gene

([R] before map means 'ruler set')

More Options:

☐ Show Connections ☒ Verbose Mode

Compress Map: Auto Compress if > px

Page Length:

Thumbnail View: ☒ default (ideogram) ☐ master

Organism: **Homo sapiens** [Help](#)

Chromosome: Region Shown:

Available Maps: **Maps Displayed (left to right):**

Org: Assembly:

–Sequence Maps–

- Ab initio
- Assembly
- BES_Clone
- Clone
- NCL_Clone
- Contig
- Component
- CpG Island

- ☒ Variation
- ☐ Gene

([R] before map means 'ruler set')

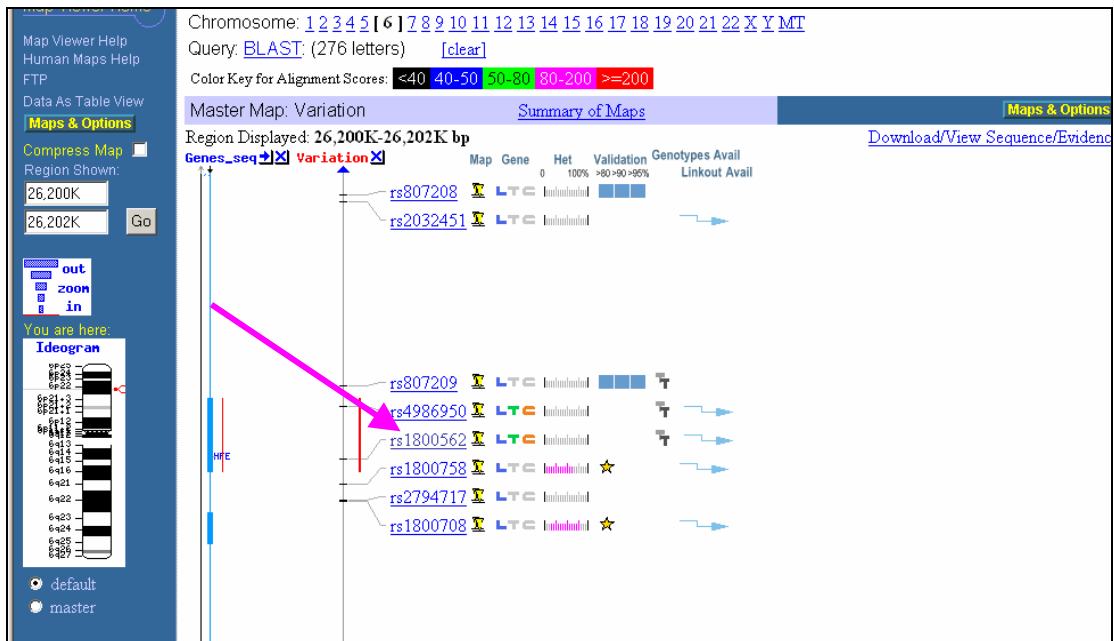
More Options:

☐ Show Connections ☒ Verbose Mode

Compress Map: Auto Compress if > px

Page Length:

Thumbnail View: ☒ default (ideogram) ☐ master



NCBI Single Nucleotide Polymorphism

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP

Search Entrez for

BUILD 126

GENERAL
 Contact Us
 dbSNP Homepage
 SNP Science Primer
 Announcements
 dbSNP Summary
 FTP Download

refSNP ID: rs1800562	Allele	Links
Organism: human (<i>Homo sapiens</i>)	Variation Class: SNP: single nucleotide polymorphism	
Molecule Type: Genomic	Alleles: A/G	
Created/Updated in build: 89/123	Ancestral Allele: Not available	
Map to Genome Build: 36.1		

SEARCH
 Entrez SNP
 Blast SNP
 Batch Query
 By Submitter
 New Batches
 Method
 Population
 Detail
 Class
 Publication
 Chromosome Report
 Locus Information
 STS Markers
 Free Form Search

Fasta sequence (Legend)

>gnl|dbSNP|rs1800562|allelePos=202|totalLen=450|taxid=9606|snpclass=1|alleles=A/G|mol=Genomic|build=113

```

ATGTGAYCTC TTCAGTGACC ACTCTACGGT GTCGGGCCTT GAACTACTAC CCCAGAAACA
TCACCATGAA GTGGCTGAAG GATAAGCAGC CAATGGATGC CAAGGAGTTC GAACCTAAAG
ACGTATTGCC CAATGGGGAT GGGACCTACC AGGGCTGGAT AACCTTGGCT GTACCCCTG
GGGAAGAGCA GAGATATACG T
R
CCAGGTGGAG CACCCAGGCC TGGATCAGCC CCTCATTGTG ATCTGGGGTA TGTGACTGAT
GAGAGCCAGG AGCTGAGAAA ATCTATTGGG GGTTRAGAGG AGTGCCTGAG GAGGTAATTA
TGCCAGTGAG ATGAGGATCT GCTCTTTGTT AGGGGGTGGG CTGAGGGTGG CAATCAAAGG
CTTTAACTTG CTTTTCTGT TTTAGAGGCC TCACCGTCTG GCACCOCTAGT CATTGGAGTC
ATCAGTGG
  
```

GeneView: no link established by BLAST analysis of mRNA sequences

Integrated Maps:

NCBI MapViewer: rs1800562 maps exactly once on NCBI human [chromosome 6](#)

Chromosome	Contig accession	Contig position	Chromosome position	Hit orientation	Contig Allele	Assembly Type	Group label	Contig label	Neighbor SNP	SNP flank position
6	NT_086686.1	25684223	25967140	plus	G	alt_assembly_2	Celera	Celera	view	201
6	NT_007592.14	16951392	26201120	plus	G	ref_haplotype	reference	reference	view	201

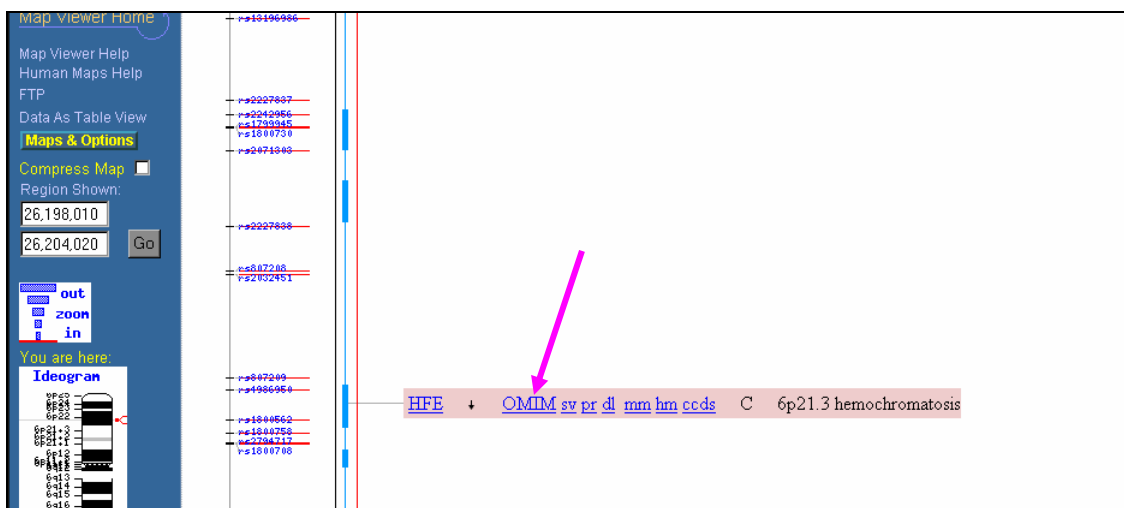
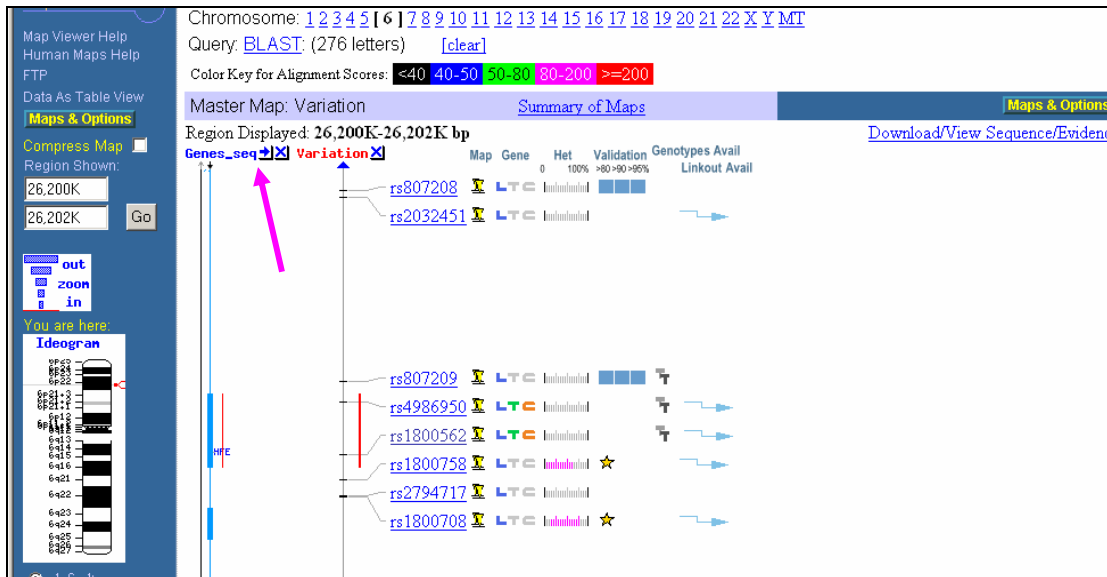
Click to see [all](#) [cDNA](#) [has frequency](#) [genomic map](#) [haplotype tagged](#) variations associated with this gene.

Group Label	Contig->mRNA	Gene Model (contig mRNA transcript) Color Legend
reference	NT_007592->NM_000410 sv function	

Group label	Contig->mRNA->Protein	Contig position	mRNA orientation	mRNA pos	Function	dbSNP allele	Protein residue	Codon pos	Amino acid pos
reference	NT_007592->NM_000410->NP_000401	16951392	forward	1066	nonsynonymous	A	Tyr [Y]	2	282
					contig reference	G	Cys [C]	2	282

Result: The EST sequence contains a known SNP (G/A with respect to the nucleotide 16951392 on contig NT_007592.14).

Step 4: Determine whether a mutant HFE gene causes a phenotype



MIM +235200

- Description
- Clinical Features
- Other Features
- Inheritance
- Mapping
- Heterogeneity
- Molecular Genetics
- Genotype/Phenotype
- Correlations
- Diagnosis
- Clinical Management
- Population Genetics
- Pathogenesis
- Cloning
- Biochemical Features
- Gene Structure
- Gene Function
- Nomenclature
- Animal Model
- History
- Allelic Variants
- View List
- See Also
- References
- Contributors
- Creation Date
- Edit History

- Clinical Synopsis
- Gene map

Online Mendelian Inheritance in Man

Johns Hopkins University

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM

Search OMIM for Go Clear

Limits Preview/Index History Clipboard Details

Display Detailed Show 20 Send to

All: 1 OMIM dbSNP: 1 OMIM UniSTS: 1

+235200 GeneTests, Links

HEMOCHROMATOSIS; HFE

Alternative titles; symbols

HLA
HEMOCHROMATOSIS, HEREDITARY; HH
HFE GENE, INCLUDED; HFE, INCLUDED

Gene map locus [6p21.3](#)

TEXT

DESCRIPTION

The clinical features of hemochromatosis include cirrhosis of the liver, diabetes, hypermelanotic pigmentation of the skin, and heart failure. Primary hepatocellular carcinoma (HCC; [114550](#)), complicating cirrhosis, is responsible for about one-third of deaths in affected homozygotes. Since hemochromatosis is a relatively easily treated disorder if diagnosed, this is a form of preventable cancer.

MIM +235200

- Description
- Clinical Features
- Other Features
- Inheritance
- Mapping
- Heterogeneity
- Molecular Genetics
- Genotype/Phenotype
- Correlations
- Diagnosis
- Clinical Management
- Population Genetics
- Pathogenesis
- Cloning
- Biochemical Features
- Gene Structure
- Gene Function
- Nomenclature
- Animal Model
- History
- Allelic Variants
- View List
- See Also
- References
- Contributors
- Creation Date
- Edit History

- Clinical Synopsis
- Gene map

Online Mendelian Inheritance in Man

Johns Hopkins University

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM

Search OMIM for Go Clear

Limits Preview/Index History Clipboard Details

Display Allelic Variants Show 20 Send to

All: 1 OMIM dbSNP: 1 OMIM UniSTS: 1

+235200 GeneTests, Links

HEMOCHROMATOSIS; HFE

ALLELIC VARIANTS
(selected examples)

- 0001 HEMOCHROMATOSIS [HFE, CYS282TYR] dbSNP
- 0002 HEMOCHROMATOSIS [HFE, HIS63ASP] dbSNP
- 0003 HEMOCHROMATOSIS [HFE, SER65CYS] dbSNP
- 0004 HFE INTRONIC POLYMORPHISM [HFE, 5569G-A]
- 0005 HFE POLYMORPHISM [HFE, VAL53MET] dbSNP
- 0006 HFE POLYMORPHISM [HFE, VAL59MET] dbSNP
- 0007 PORPHYRIA VARIEGATA [HFE, GLN127HIS] dbSNP
- 0008 HEMOCHROMATOSIS [HFE, ARG330MET]
- 0009 HEMOCHROMATOSIS [HFE, ILE105THR] dbSNP
- 0010 HEMOCHROMATOSIS [HFE, GLY93ARG] dbSNP
- 0011 HEMOCHROMATOSIS [HFE, GLN283PRO]

Result: Mutations in the HFE gene are associated with hemochromatosis disease.

Problem 2:

<http://www.ncbi.nlm.nih.gov/Class/minicourses/diseasegene2.html>

A laboratory has generated an EST library from a sickle cell anemia patient and wants to identify the gene(s) causing the phenotype. Sickle cell anemia is a disease in which the red blood cells are curved in shape, and which causes pain and fever.

Outline:

We will follow these steps to solve the problem:

1. Compare ESTs from a sickle cell anemia patient to the human genome (using BLAST).
2. Identify the gene(s) aligning the ESTs and download their sequences (using Map Viewer).
3. Identify whether the ESTs contain any known nucleotide variations (single nucleotide polymorphisms) (using dbSNP).
4. Determine whether a mutant form of the gene is known to cause a phenotype (using OMIM).

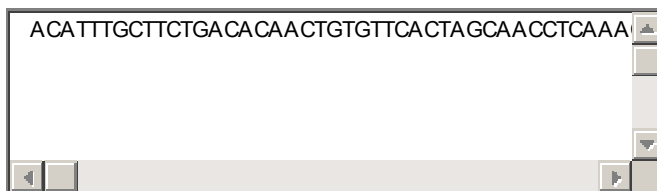
Step 1. Compare ESTs to the human genome (using BLAST):

One way to identify the genes expressing the ESTs is to compare their sequences using BLAST with the human genome assembly and the genes annotated on it. To access the specialized BLAST page for searching against the human genome assembly, click on

BLAST (human genome)

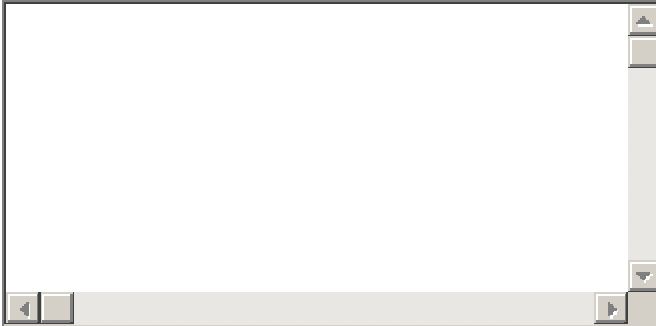
Paste the EST sequence provided below in the query box of the BLAST page and start the search by clicking on the “Begin Search” button.

Query EST Sequence:



Name the chromosome and the contig that we get as a BLAST hit. Note that the similarity is on the minus strand of genome. Is the EST sequence 100% identical to the genomic sequence? Note the nucleotide difference between the two sequences. Paste your results in the window below.

Results of BLAST against the human genome

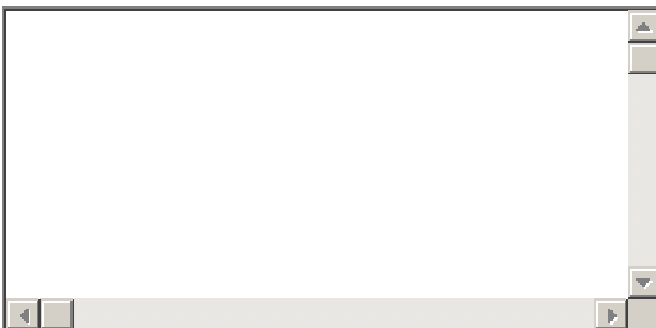


Step 2. Identify the gene(s) expressing the ESTs and download their sequences:

To visualize the BLAST hit on the genome using Map Viewer, click on the "Genome View" button at the top of the results page, then on the Map element "NT_009237". Currently, 4 maps should be displayed (Contig, Model, RNA and Gene_seq). Zoom out 2 or 4 times by clicking on right most contig map and selecting the appropriate option.

The best BLAST hits, indicated by the red bars, are in the region of two exons of the HBB gene annotated on the human genome. Make the Gene_seq map a master map by clicking on the arrow at the top of the map. Note that the gene is annotated on the minus strand. To display the entire HBB gene sequence, click on the "dl" link, choose minus strand from the pull down menu, click on "Change Region/Strand" and display the sequence by clicking on "Display". Copy the sequence and paste it in the area provided below. We will use it later to obtain the exon-intron structure. You can adjust the nucleotide locations to download the upstream or downstream sequence by using the "adjust by" and "Change Region/Strand" option.

HBB gene sequence



Step 3. Determine whether the ESTs contain known SNPs:

Go back to the Map Viewer report. Click on the Maps and Options link. Remove all the maps except the Gene_seq map by selecting the map under the Maps Displayed menu and clicking on Remove. Now add the variation map from the Available maps menu (by selecting the map and clicking on Add). Make the Variation map as the master map by selecting it and clicking the Make Master/Move to Bottom option. Then click on Apply. Now two maps are displayed, Variation (it's the rightmost and the master map) and Gene_seq. The master map provides detailed information for the map features, in this case SNPs. ". (The Mini-Course Map Viewer Quick Start describes the usage of the Map Viewer in detail.) Zoom in on the blast hit area (red bar). There are two SNPs in the area, one of them is rs334. Click on the link for the SNP. There is an A/T SNP is at the nucleotide position 4035473 on the contig NT_009237 as mentioned under Fasta sequence and Integrated maps. Is this the same nucleotide variation found in the BLAST result in Step 1?

Step 4. Determine whether the mutant HBB gene causes a phenotype:

Go back to the Map Viewer report. Make the Gene_seq map as the master map. Select the link to the OMIM database. It takes us to the OMIM report for the HBB gene that details how mutations in the HBB gene are associated with a phenotype, sickle cell anemia. As mentioned in the report, the allelic variants are listed for the mature HBB protein which lacks initiator methionine. Click on the Allelic Variant "View list" to get information about mutant proteins from patients. Is Glu6Val variant mentioned in the list? Which phenotype does it cause?

Summary:

This mini-course describes steps to identify the gene expressing the ESTs obtained from a sickle cell anemia patient, download the gene sequence, identify known SNPs in the gene and find SNP-associated phenotypes.

Step 1: The query EST sequence was found to align contig NT_009237.17 on chromosome 11 with one nucleotide difference (T to A with respect to the nucleotide 4035473 on the contig).

Step 2: The query EST was found to be expressed by the HBB gene.

Step 3: The query EST sequence contains a known SNP (T/A with respect to the nucleotide 4035473 on contig NT_009237.17).

Step 4: Mutations in the HBB gene are associated with sickle cell anemia.